

Public Emotional Diffusion over COVID-19 Related Tweets Posted by Major Public Health Agencies in the United States

Haixu Xi^{1,2}, Chengzhi Zhang^{1†}, Yi Zhao¹ & Sheng He²

¹School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China

²School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China

Keywords: Emotional diffusion; Tweets; COVID-19; Pandemic management; US Public Health Agency

Citation: Xi, H.X.: Public emotional diffusion over COVID-19 related tweets posted by major public health agencies in the United States. *Data Intelligence* 4(1), 66-87 (2022). doi: 10.1162/dint_a_00101

Received: February 23, 2021; Revised: March 25, 2021; Accepted: May 18, 2021

ABSTRACT

Since the end of 2019, the COVID-19 outbreak worldwide has not only presented challenges for government agencies in addressing public health emergency, but also tested their capacity in dealing with public opinion on social media and responding to social emergencies. To understand the impact of COVID-19 related tweets posted by the major public health agencies in the United States on public emotion, this paper studied public emotional diffusion in the tweets network, including its process and characteristics, by taking Twitter users of four official public health systems in the United States as an example. We extracted the interactions between tweets in the COVID-19-TweetIds data set and drew the tweets diffusion network. We proposed a method to measure the characteristics of the emotional diffusion network, with which we analyzed the changes of the public emotional intensity and the proportion of emotional polarity, investigated the emotional influence of key nodes and users, and the emotional diffusion of tweets at different tweeting time, tweet topics and the tweet posting agencies. The results show that the emotional polarity of tweets has changed from negative to positive with the improvement of pandemic management measures. The public's emotional polarity on pandemic related topics tends to be negative, and the emotional intensity of management measures such as pandemic medical services turn from positive to negative to the greatest extent, while the emotional intensity of pandemic related knowledge changes the most. The tweets posted by the Centers for Disease Control and Prevention and the Food and Drug Administration of the United States have a broad impact on public emotions, and the emotional spread of tweets' polarity eventually forms a very close proportion of opposite emotions.

[†] Corresponding author: Chengzhi Zhang (Email: zhangcz@njust.edu.cn; ORCID: 0000-0001-9522-2914).

1. INTRODUCTION

Since the end of 2019, the global outbreak of COVID-19 has caused plenty of political and social issues. Governments of various countries have successively introduced measures to respond to the COVID-19 pandemic and shared news and information through a variety of information channels. Twitter is one of the most popular social media platforms, and tweets reveal the government's progress in fighting the pandemic more in real time. Researchers were interested in the emotional analysis of different types of tweets published by government agencies and attempted to investigate the public's attitudes about pandemic prevention policies and measures taken by governments [1, 2]. Also, studies were conducted in the diffusion law of public information, such as the diffusion of academic results [3], conspiracy [4], topics about COVID-19 [5] and emotion analysis through social networks [6]. The diffusion speed of emotional information is faster and more active in comparison with other types of information [7], which implies that research on emotional diffusion of the public over COVID-19 related tweets posted by government agencies may shed light on the trend and pattern of opinion changes of the public on the pandemic management conducted by a government. Meanwhile, through the research, we can try to identify key users that affect attitudes of others to grasp the group polarization phenomenon of the public, and further provide reference for government agencies to avoid unreasonable policies against COVID-19 pandemic and provide insights into public opinion management.

Nevertheless, there is insufficient research into the emotional diffusion of the public over COVID-19 related tweets posted by government agencies. Meanwhile we are confronted with many difficulties and challenges in studying emotional diffusion of the public opinion. Zafarani et al. [8] have identified challenges in the measurement of the characteristics of emotional diffusion and the analysis of users with different roles. Until now, various methods and technologies have been applied in analyzing emotional diffusion in different domains but there still have some issues to be solved such as an analytical framework to consider the influence of different factors on the mode of emotional communication in different situations.

This paper aims to propose a method for characterizing the emotional diffusion network of COVID-19 related tweets posted by US government agencies and measuring its features, which supports dynamic visualization of emotional diffusion process of tweets. Meanwhile a method of analyzing the role of key users is also proposed in the process of observing emotional diffusion and the emotional influence on subsequent users from the perspective of change of emotional intensity and the proportion of emotional polarity. In order to detect the characteristics of emotional diffusion under different influencing factors, this paper also takes the different characteristics of tweets published at different time periods, tweet topics and publishing agencies into consideration.

2. RELATED WORK

The emotional diffusion of the public is defined as dissemination of the emotional expression characteristics [9], such as emotional intensity and polarity of the public. There are some systems of analyzing emotional diffusion proposed in the literature. For instance, Trung et al. [10] developed the

chinaXiv:202211.00379v1

TweetScope system based on fuzzy propagation models for emotional analysis on online social network. In addition, some researchers have conducted statistical and correlation analysis on public sentiment and its diffusion indicators, and summarized the basic characteristics and laws of emotional diffusion. Xu [11] found that there is a higher degree of positive correlation between the popularity of news dissemination and anger than its association with other emotions. In order to understand the complex network characteristics of emotional diffusion, most researchers first used the social network analysis method to construct the network structure, and then analyzed the distribution characteristics of the network. For example, Miller et al. [12] found the rule of emotional diffusion based on the characteristics of cascade network of sentiment. In order to study the formation mechanism of emotional diffusion, researchers built a mathematical calculation model to predict the emotional changes. Using the independent cascade model of sentiment, Xiong et al. [13] introduced the measure of personal sentiment transitivity and found the emotional diffusion in heterogeneous social media.

Various factors will affect the emotional diffusion process. Users with different characteristics and influence, different emotions and different event types have different characteristics of emotional diffusion, diffusion mode and influence [14]. The existing research does not pay enough attention to the modes of emotion diffusion, and the influencing factors of emotional diffusion combined with different event situations. Previous research mainly focused on the emotion diffusion between interactive users. In this paper we attempt to focus on the public emotion implied in tweets and study the structure, process and characteristics of the diffusion network of emotion between interactive tweets, in an effort to propose an analytical framework to investigate public emotional diffusion of tweets related to the COVID-19 pandemic and verify its effect on the change of public emotion responses.

This paper aims to solve the following two problems: First, what is the impact of official tweets from major agencies of the public health system on public sentiment. Second, how does the public's emotion spread in the process of tweeting, commenting and mentioning, and what are the characteristics and rules of the public's emotion. Whether there are differences in public emotional diffusion in tweets published in different pandemic stages, different tweet topics and different tweet posting institutions.

3. METHODOLOGY

The US government plays its role in pandemic management through the public health system, while the US National Institutes of Health (NIH), Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC) are the core of this system [15]. The Department of Health and Human Services (HHS) is in charge of the aforementioned institutions directly. These official agencies are the authoritative channels for the American people to obtain information about the pandemic. Their tweets during the pandemic directly reflect the measures of the United States to deal with the pandemic.

As shown in Figure 1, this research process mainly includes three parts. First, we extracted tweets published by HHS, NIH, CDC and FDA and their public interactive tweets, including tweet data and behavior data from the open-source data set COVID-19-TweetIds, and then the tweet data were clean and

tokenized. Second we extracted the interaction between tweets from the behavior data, extracted the topic of tweets from the preprocessed data, calculated the emotional intensity of tweets and determined the corresponding emotional polarity. These three parts of data were integrated to the required corpus^① for research.

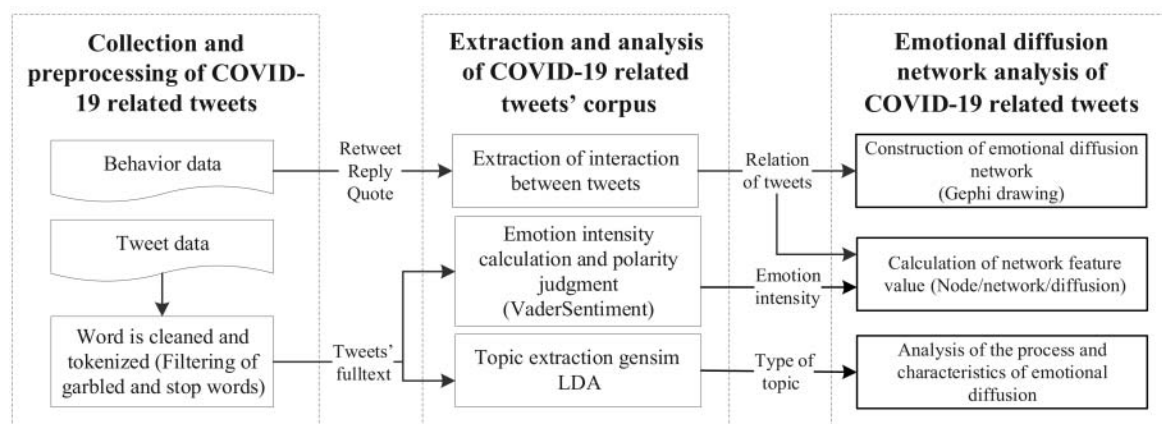


Figure 1. The research framework of public emotional diffusion of tweets related to the COVID-19 pandemic.

In the third part, this paper put forward the method of constructing the emotional diffusion network, the measurement method of network characteristics and the analysis method of emotional diffusion process based on the interaction data in the analysis of emotional diffusion network of COVID-19 related tweets. This paper first analyzed the changes of public emotions in the process of emotional diffusion of official tweets and the changes of public emotions caused by key nodes or users. Then, the differences of emotional diffusion characteristics of tweets on different publishing dates, in topic categories and by different posting departments were studied to summarize their diffusion characteristics.

3.1 Collection and Preprocessing of COVID-19 Related Tweets

In this study, we used the continuously updated open source data set: COVID-19-TweetIds [16] provided by Emily, and programming Python script to obtain more than 129 million COVID-19 related tweets with *twarc* component^② according to the twitter ID, with the time range from January 21, 2020 to May 31, 2020. Then, 356 tweets closely related to COVID-19 were screened out, including 141 source tweets, accounting for 39.61%, and 215 forward and reply tweets, accounting for 60.39%.

Data preprocessing is the basis of the construction and analysis of the structure of an emotional diffusion network, mainly including data cleaning and word tokenized. This paper filtered and deleted special

^① All the corpus is available the Science Data Bank repository, <https://doi.org/10.11922/sciencedb.01044>, under an Attribution 4.0 International (CC BY 4.0).

^② <https://github.com/DocNow/twarc>

characters, garbled codes, hyperlinks and special signs, stop words in tweets, and converted tweets to lowercase form. A user-defined stop word list was used, which was based on the general English stop words in NLTK toolkit^③. The meaningless words of the high-frequency words were added into the list by manual filtering. Word tokenized refers to the process of recombining continuous character sequences into word sequences according to certain norms. This study used space, punctuation and other markers to segment tweets.

3.2 Extraction and Analysis of COVID-19 Related Tweets' Corpus

The emotional intensity of tweets and the topics they belong to are an important part of the emotional diffusion network structure, which can reflect the evolution of topics that users pay attention to and emotional changes in the diffusion network, and the interaction between tweets is the basis for building an emotional diffusion network.

3.2.1 Extraction of interaction between COVID-19 Related Tweets

The interaction between tweet nodes is represented by $Relation_{AB}$, and its information structure is as follows:

$$Relation_{AB} \{Node\ A\ tweetId, Node\ B\ tweetId, level, weight\}$$

where $Relation_{AB}$ represents the interaction between tweet Node A and Node B, Node A *tweetId* is the tweet ID of the parent node and Node B *tweetId* is tweet ID of the child node, and *level* represents the current diffusion level of the interaction. There are three types of interaction between tweets: directly forwards, reply and quotation. *Weight* is used to measure the degree of interaction. It is generally believed that direct forwarding is a simple concern, while reply means paying more attention to the parent node, and reference means recommending to others at the same time, indicating the most attention. The *weight* value for the directly forwards is 1, and for reply and quotation are 2 and 3, respectively.

The pseudo code for extracting the interaction between tweet nodes is shown in Algorithm 1. First, the ID, reply ID, forwarding ID and quotation ID of each tweet from the full tweet data set were extracted, saved as a json file and imported into the collection of *Mongodb*^④. Then the list of high-profile tweets was read, and the database collection was scanned to find the association interaction of each tweet in the list.

The subroutine of finding the interaction between tweets is to read a tweet's ID in the list of high-profile tweets at first, and search for the tweet IDs list with that ID in the set of reply ID, forwarding ID, or quotation ID. If it exists, the interactions between the tweet's ID and the tweet IDs found is recorded circularly, and then the tweet IDs found is used as the new ID, respectively, and the list of associated tweet IDs with this ID is continued to be searched in the set, and the loop is repeated until all interactions are found.

^③ <https://www.nltk.org/>

^④ <https://www.mongodb.com/>

Algorithm 1: Relation extraction between tweets nodes.**Input:** The tweet data set: COVID-19-TweetIDs, T_n ; Tweet list with high degree of concern, L_n ;**Output:** The relation record between tweets, R_n ;1: Read tweets' reply Id, forward Id and quotation Id from T_n ; **while** not end of L_n **do** compute $Jaro(r_i, r_m) (r_m \in T_j)$; **if** $(Jaro(r_i, r_m) \geq \theta_r) \wedge (Jaro(r_i, r_m) \geq r^{maxSim})$ **then** replace r^{maxSim} with r_m ;2: $E_n = E_{n-1} \cup E$;3: Classifying samples in $U_n - T_n$ by E_n ;4: Deleting some weak classifiers in E_n so as to keep the capacity of E_n ;5: **return** E_n ;

3.2.2 Calculation of Emotional Intensity for COVID-19 Related Tweets

The calculation of emotional intensity adopted the popular emotional analysis tool *vaderSentiment*[®]. It is based on a manually annotated dictionary, which contains tens of thousands of words, punctuation marks, network expressions, emoticons and corresponding emotional intensity and polarity. Before calculating the emotional intensity of a sentence, the sentence structure is first regularized according to the grammar rules, then the emotional intensity index of each word is searched according to the dictionary, and finally the emotional intensity of the sentence is combined and calculated. The effectiveness outperforms 11 typical state-of-practice benchmarks including LIWC, ANEW, the General Inquirer, SentiWordNet, and machine learning oriented techniques relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms[17]. In this paper, we utilized this component to calculate the emotional *intensity* of the preprocessed tweet, and to determine the emotional *polarity* of the tweet with Equation (1).

$$Polarity = \begin{cases} Negative, & -1 \leq Intensity \leq -0.05 \\ Neutral, & -0.05 < Intensity < 0.05 \\ Positive, & 0.05 \leq Intensity \leq 1 \end{cases} \quad (1)$$

3.2.3 Topic Extraction of COVID-19 Related Tweets

Commonly used words or phrases are always implied in a topic, and latent Dirichlet allocation (LDA) is used to extract topics of tweets. Most of the tweets are short texts, and research shows that the LDA model is not very effective in topic extraction on short texts. Considering that the replies, reposts and comments of a tweet have a high probability of being similar to the topic of the tweet, this research first merged a tweet with all its replies, forwards and comments, preprocessed it, then used the NLTK-Rake to extract the phrases in the tweets, set the number of topics and the number of words (or phrases) under each topic and then used the LDA module in the *gensim*[®] component to obtain the results of the topic model. Each of the

[®] <https://github.com/cjhutto/vaderSentiment>[®] <https://radimrehurek.com/gensim/>

words (or phrases) under the topic were summarized to determine the name of the topic. Finally, the trained model was used to predict the topic to which the tweet belonged. The following three themes were summarized: pandemic prevention management measures, pandemic related knowledge, and alert of pandemic progress with the keywords and key phrases returned by the model.

3.3 Emotional Diffusion Network Analysis of COVID-19 Related Tweets

3.3.1 Feature Measurement of Emotional Diffusion Network

The characteristic value of the network structure of emotional diffusion is similar to that of the traditional social network structure. There are three types of node attributes, network attributes and propagation attributes. Node attributes are characterized by node centrality, indicating the value and influence of nodes in the network, which mainly include relative degree centrality, relative proximity centrality, and relative betweenness centrality [18]. As shown in Equation (2), $C_{btw}(v)$ is the relative betweenness centrality of node v , which measures the mediating effect of the node on the spread of network emotion. This article analyzed the changes of emotional intensity of key nodes (nodes with higher value of the relative betweenness centrality) and key users (users corresponding to this node) to reveal the role of intermediary nodes in the process of emotional diffusion.

$$C_{btw}(v) = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \quad (2)$$

where N is the set of all nodes in the network, $\sigma_{s,t}$ is the number of the shortest path between node s and node t , and $\sigma_{s,t}(v)$ is the number of the shortest path passing through node v between node s and node t .

The network attributes indicate the overall situation of the network, including the number of nodes, links and the density, the radius/diameter, the average shortest distance of the network. The network radius is the smallest node eccentricity, the network diameter is the largest node eccentricity, and the node eccentricity is the maximum value of the distance between a node and all other nodes in the network. As shown in Equation (3), D_{net} is the diameter of the network, and $C_{ecc}(N)$ is the eccentricity of all nodes in the network.

$$D_{net} = \text{Max}(C_{ecc}(N)) \quad (3)$$

The spread attribute indicates the influence of the tweet spreading process, including the extent (CS_{net}), depth (DS_{net}) and speed (VS_{net}) of diffusion. Among them, the extent of diffusion is the sum of the out-degrees of the nodes under the tweet node, and the depth of diffusion is the eccentricity of the source node. The diffusion rate is shown in Equation (4), where TS_{net} is the diffusion time, the unit can be seconds, and C is the coefficient, which can take any integer, such as 10,000.

$$VS_{net} = C \times \frac{CS_{net}}{TS_{net}} \quad (4)$$

3.3.2 Construction of the Emotional Diffusion Network of COVID-19 Related Tweets

In the construction of the emotional diffusion network, each tweet is regarded as a network node, the connection between nodes represents the interaction between nodes, and the weight represents the type of interaction between nodes.

(1) Information composition of the node

The diffusion network structure of tweets is mainly implied in information of reposting, quotation and reply of tweets. The fields beginning with *in_reply_to_status* store the information about the original tweet that this tweet replied to. Reposting has two types: direct reposting and reposting with comments (also called citations). The *retweeted_status* field stores the relevant information of the original tweet directly reposted by this tweet. This tweet only adds *RT* in front of the original tweet content. The emotion and theme are the same as the original tweet. The *quoted_status* field stores the relevant information of the original tweet quoted by this tweet.

The information composition of the tweet node is as follows:

$Node A_{tweetinfo}(tweeted) \{tweeted, created_time, userid, senti_score, topic\}$

$Node A_{userinfo}(userid) \{userid, u_name\}$

$Relation_{AB} \{Node A_tweetId, Node B_tweetId, level, weight\}$

where $Node A_{tweetinfo}$ represents the basic information of node A, *senti_score* is the emotional strength of the tweet, *topic* is the topic to which the tweet belongs, and *u_id* and *u_name* are the tweet user ID and username, respectively. $Relation_{AB}$ represents the interaction between tweet nodes.

(2) Drawing of network structure

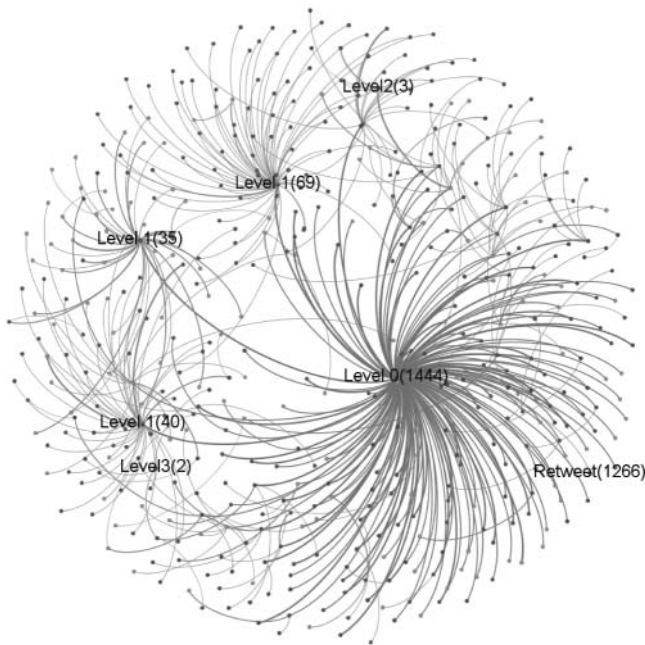
Emotional diffusion network includes one-to-one and one-to-many relationships. This study uses Gephi® to visualize the emotional diffusion network according to the structure information of nodes. Each tweet is mapped to a node within the network, and the interaction between tweets is mapped to the edge between nodes. The thickness and color of the edge represent the type of interaction between the nodes and the diffusion level, respectively, and the label description of the node can be the sum of the number of all interaction nodes under the current node, or the emotional strength of the current node. The color of the node indicates the emotional polarity of the current node.

(3) Case analysis: Taking the knowledge popularization of COVID-19 published by CDC as an example

This paper shows the emotional diffusion network and its characteristic value by choosing the tweet that ranks first in the total number of forwarding and likes as an example. The tweet ID is “1220829014811607043”, released by CDC, is knowledge about the spread of COVID-19 virus and symptoms. As shown in Figure 2, the network propagates four levels (level 0 - Level 3), of which the two tweets in layer 4 are all direct

® <https://gephi.org/>

forwarding tweets. There are 1,266 tweets directly forwarded in the first level. In the whole process of emotional transmission, the proportion of positive emotion was 32.33%, the proportion of neutral emotion was 32.76%, and the proportion of negative emotion was 34.91%.



Nodes	1,749	Edges	1,748
Diameter of Network			8
Scope of Spread (n)			304
Degree of Spread (l)			4
Velocity of Spread (n/s)			0.52

Figure 2. Tweets’ emotional diffusion network and their characteristic values. Because the number of tweets forwarded at the first level is huge, and most of them are directly forwarded, in order to optimize the drawing effect of the sentiment diffusion network, the tweets directly forwarded at the first level are simplified, and only the tweets with the earliest direct forwarding time are retained.

3.3.3 Analysis of the Process and Characteristics of Emotional Diffusion of COVID-19 Related Tweets

In the process of forwarding, replying and quoting from different levels of users, the emotion of the source tweet has spread, and its intensity will change to a certain extent. It is also possible that some nodes have mutation and emotion reversal. The nodes whose relative betweenness centrality exceeds the average value are called key nodes, and the users to which these nodes belong are called key users. Therefore, the process of the emotional diffusion network can be described from the number of emotional intensity and polarity changes of different diffusion levels. The emotional changes after the key nodes in the process of diffusion and the key user’s emotional changes also need to be analyzed.

In this paper, the interaction tweets of COVID-19 related tweets released by the major public health organizations in the United States were collected and counted according to the diffusion level. The average intensity of emotion and the average proportion of different emotional polarity levels were calculated, and then the average intensity of the emotional diffusion network and the change chart of different emotional polarity levels’ average proportion were plotted, respectively. Finally, we analyzed the average proportion

of different emotional polarity levels after the key nodes of the emotional diffusion network and the change graph of the average emotional intensity after the key users. In order to analyze the characteristics of the emotional diffusion network, this paper compared the emotional diffusion network of tweets in different release months, under different topic categories and released by different departments, and drew the average emotional intensity and the average proportion of emotional polarity from the dimension of diffusion level.

4. RESULTS

This paper described the statistical distribution of the characteristics of the emotional diffusion network of four official twitters in the public health system in USA, and analyzed the dynamic propagation process of all the above tweets. We found that the characteristics of the emotional diffusion of the public to the relevant tweets varied with different release months, topic categories and release agencies.

4.1 Descriptive Statistics of Emotional Diffusion of COVID-19 Related Tweets

As shown in Figure 3, the date when tweets were created is mainly from January to March, which is also the peak period of the outbreak of COVID-19 in the world; the polarity of tweets tends to be positive and neutral, accounting for a relatively high proportion; the topic of tweets mainly includes pandemic prevention management measures, pandemic related knowledge and alert of pandemic progress. And pandemic prevention management measures include virus detection, vaccine research, clinical treatment, material procurement and community management. The total number of source tweets published by CDC and HHS is ranked first and second, respectively. Although the extracted tweets are a subset of the complete set of COVID-19 tweets, the distribution of tweets has obvious characteristics with a slight difference.

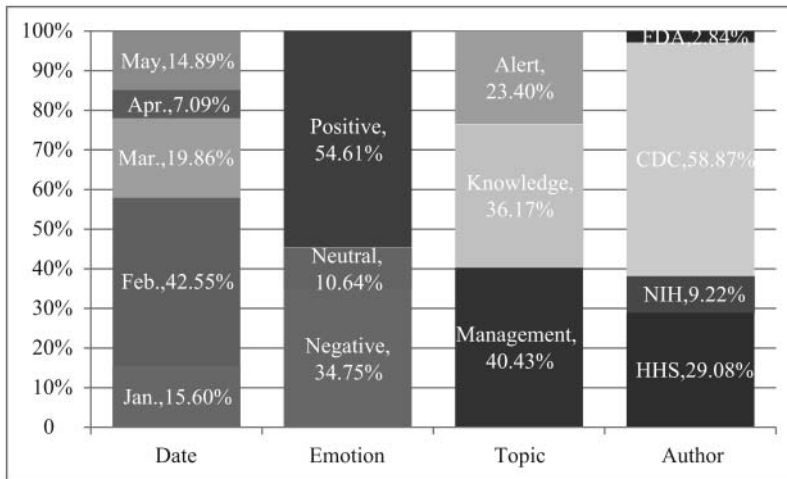


Figure 3. The distribution of tweets sent by the US public health system by month, emotional intensity, topic category, and posters. Management refers to pandemic prevention management measures, and its topic words are management, launched, press conference, COVID-19 test, and community interventions; Knowledge refers to pandemic related knowledge and its topic words include symptoms, question, answer, watch video, and need to know; Alert refers to alert of pandemic progress and its topic words are cases, latest, reports, updated, and confirm.

Table 1 shows the descriptive statistical characteristics of related attributes of tweets' emotional diffusion network from four government agencies. Although the emotional intensity of source tweet nodes shows skew distribution, in the process of emotional transmission, the distribution of the proportion of different emotional polarity levels conforms to the normal distribution. The results show that the degree of skewness of diffusion nodes, number of edges, diffusion width and relative betweenness centrality of different tweets are large, especially for the degree of skewness of the diffusion speed. Obviously, the tweets about attentions in daily life and the government's response to COVID-19 spread at the highest speed, such as "What are five things you need to know about novel" and "Today, FDA issued an EUA for CDC diagnostic to detect 2019nCoV".

Table 1. Descriptive statistical characteristics of tweet attributes.

Variable	Avg.	Median	Mode	Std.	Min	Max	Skewness
Emotional intensity of Node	0.284	0.388	-0.128	0.361	-0.480	0.856	-0.423
The proportion of positive emotions in diffusion	0.366	0.331	0	0.222	0	1.000	1.019
The proportion of neutral emotions in diffusion	0.348	0.368	0	0.211	0	1.000	0.862
The proportion of negative emotions in diffusion	0.281	0.289	0	0.220	0	1.000	1.449
diffusion level	2.36	2.00	2	1.40	1	6	1.227
Nodes	406.14	185.00	151	508.5	12	1749	1.853
Edges	405.14	184.00	150	508.5	11	1748	1.853
Extent of diffusion	65.68	5.50	0	111.37	0	334	1.584
Speed of diffusion	2306.26	0.463	0.0	6308.16	0	21000	2.276
Key nodes	1.50	1	1.439	1.44	0	5	1.001
Relative betweenness centrality of node	24.77	0	46.046	46.05	0	154	2.007

4.2 Emotional Diffusion Process of COVID-19 Related Tweets

Based on the calculation of the emotional intensity and polarity of each level of interactive tweets mentioned in the previous section, this paper analyzed the process of the emotional diffusion network of four government agencies' source tweets as a whole.

(1) The dynamic diffusion process of public sentiment

As shown in Figure 4 (a), in the process of emotional transmission of tweets, source tweets gradually turned from positive to negative emotion in the first four levels and become positive in the fifth level. The transformation from negative to positive emotion is mainly due to the response of node "1233891883195211780" on the fifth layer—"don't worry, CDC's got it", and other nodes releasing the latest meeting news and medical preparation of the government, which reversed the spread of negative among the public.

As shown in Figure 4 (b), in the process of emotional diffusion of tweets, the proportion of positive and neutral emotion fluctuates at different diffusion levels, but the proportion of negative emotion gradually increases. It shows that in the process of diffusion, the public emotion sometimes showed positive and optimistic, sometimes tended to be rational. With the further development of relevant discussions, the neutral emotion disappeared, and the public emotion finally formed a situation of differentiation and opposition.

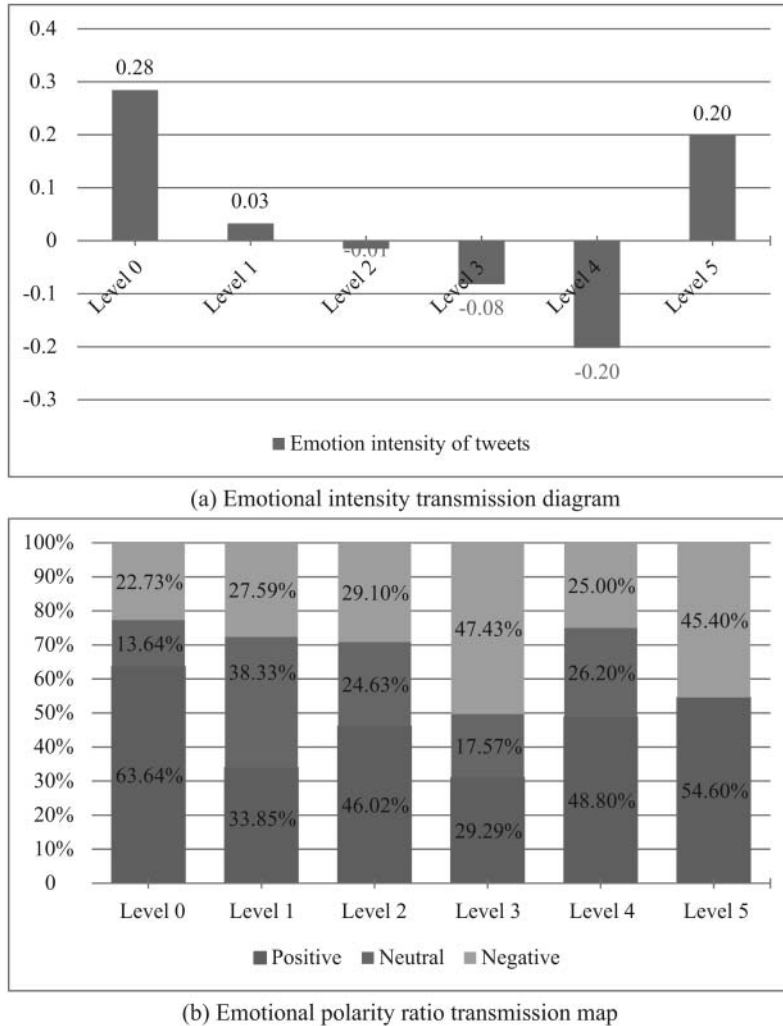


Figure 4. Dynamic spread of emotion communication network.

(2) Emotional influence process of key nodes in the diffusion network

As shown in Figure 5 (a), in the diffusion process of key nodes, negative emotions account for the majority. In general, the public does not agree with relevant topics derived from COVID-19 related tweets of the main public health institutions. For example, the node “122150049444462592” is about the CDC being the tweets of the authoritative information sources of the pandemic situation. The “12223228043742 4128” node is the tweets for the coordination work of the National Security Council, and the node “122191290149519769” is based on the *Pandemic and All hazards preparation and promoting innovation act*. As shown in Figure 5 (b), the average emotional intensity of key users is positive, but the average emotional intensity of subsequent node users of these users gradually becomes negative in the process

of diffusion. For example, the neutral emotion of node “12215004944625920” corresponding to user “160946337” gradually becomes negative in the process of transmission, which indicates that the public doubted the authenticity of CDC pandemic information. Positive emotions of the node “1221912901495197698” corresponding to the user “21157904” gradually turned negative in the process of diffusion, indicating that the public were generally skeptical about the good effect of the bill.

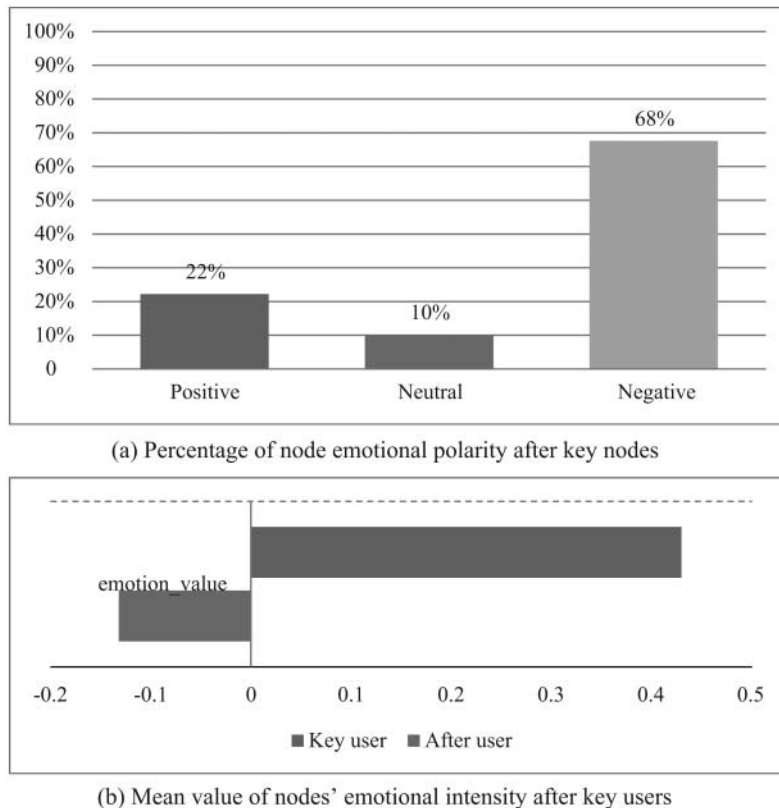


Figure 5. Emotional changes after key nodes.

The data set of COVID-19 tweets contains only part of the data, and the data will not be updated automatically, which can only reflect the emotional diffusion network of tweets observed at a certain point in timeline, and the network characteristic value calculated can only reflect the network attributes at that point in timeline, and the whole emotion diffusion network of tweets will evolve with time dynamically.

4.3 Characteristics of Emotional Diffusion Network of COVID-19 Related Tweets

(1) The differences of emotional diffusion in different months

As shown in Figure 6 (a), most of the tweets released by US public health agencies in February were reports on the progress of the pandemic in China. In March, US began to make a comprehensive report of domestic cases, and actively dealt with the epidemic, such as purchase of ventilators, and recommendation of keeping physical distance. Therefore, February and March saw the highest level of tweets, which were generally concerned by the public. In January, most of the tweets released by US public health agencies only reported the progress of the pandemic situation in China and a small number of domestic cases. In February, they also released soothing tweets, such as no suggestion to wear masks and no community infection. In March and April, the pandemic situation in US became severe, and the government took corresponding measures to encourage wearing masks and adopted strict community management which was reflected in the process of spreading tweets from January to April, and the emotional intensity gradually changed from positive to neutral or negative. With the full implementation of the government's emergency measures, the emotional intensity of tweets eventually tended to be positive and neutral. However, in February, the most serious international epidemic, negative emotions continued to spread among the public. The negative emotion in May directly turned into positive emotion, which shows that the government's response measures were improved and effective (the progress of vaccine research and medical treatment had been announced since May, and community support services were provided).

As shown in Figure 6 (b), the trend of the proportion of emotional polarity in the process of diffusion from January to May is basically consistent with the trend of emotional intensity. In February, the proportion of negative emotions increased significantly. In March, the proportion of negative emotion continued to increase, but it was mostly positive at the last level. In April, the proportion of neutral emotions increased significantly, while the proportion of positive emotion rose significantly after May. February was the outbreak time of global pandemic and public emotion tended to be negative or neutral. From March to May, the US government's pandemic prevention measures achieved certain results, and the public emotion obviously turned to be positive and neutral.

(2) The differences of emotional diffusion in tweets of different discussion topics

As shown in Figure 7 (a), the diffusion levels of all topics are relatively same, reaching 5-6 levels. Among them, in the process of emotional diffusion of tweets on *Pandemic related knowledge*, the intensity of public emotion changed from positive to negative, with the largest change range. The reason is that in April, the government began to encourage wearing masks in the tweet on *Pandemic related knowledge* and most of the public replies mentioned the government's proposal not to wear masks in February. In the topic of *Alert of pandemic*, the negative emotion continued to spread; in the topic of *Pandemic prevention management measures*, the negative emotion continued to spread, the degree of public emotional intensity changing from positive to negative was the largest, and it happened at a lower level.

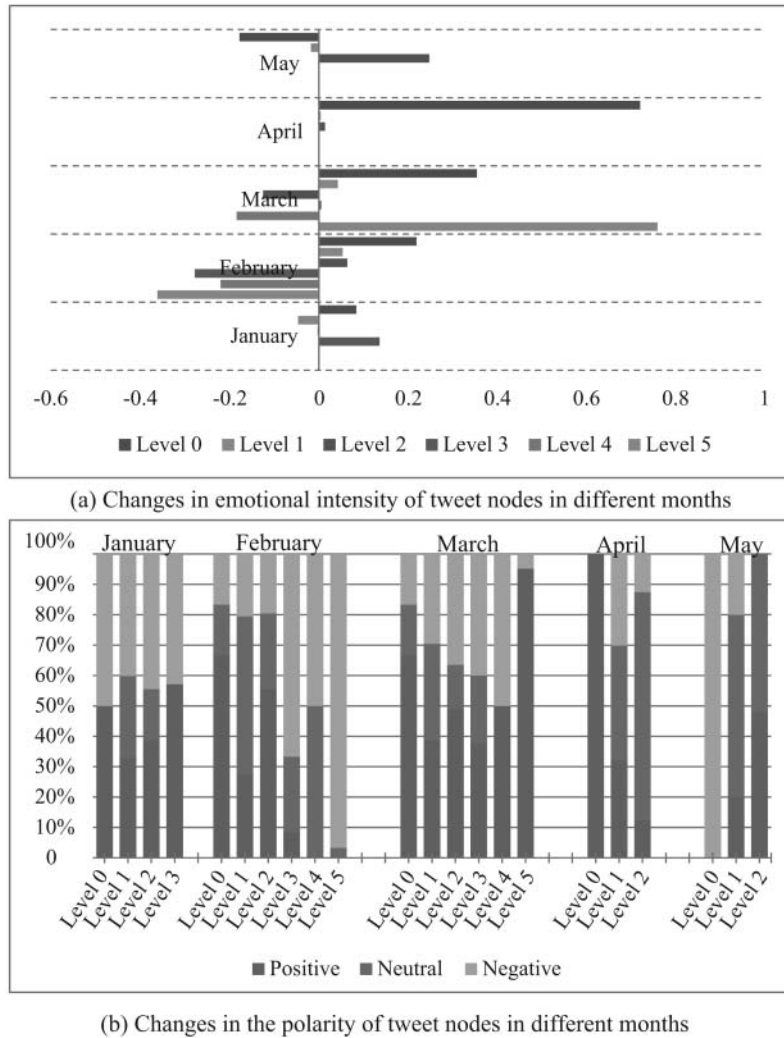


Figure 6. The emotional diffusion of tweet nodes in different months.

As shown in Figure 7 (b), the positive and neutral emotions on the topic of *Pandemic Related knowledge* gradually decreased, while the negative emotions increased significantly. The negative emotion on the topic of *Alert of pandemic* gradually decreased, but gradually increased to the highest proportion in the end. The results show that the positive emotion on the topic of *Pandemic prevention management measures* gradually decreased, the negative emotion gradually increased, and the neutral emotion accounted for the highest proportion.

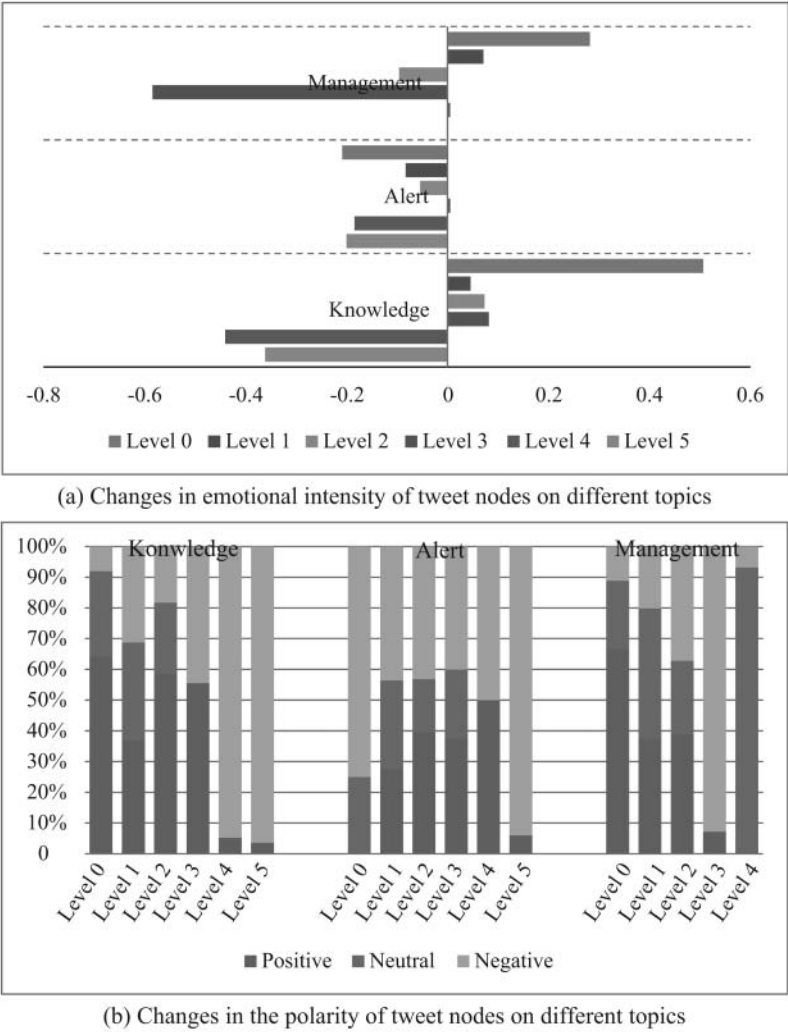


Figure 7. Emotion spread of tweet nodes on different topics.

(3) The differences of emotional diffusion in tweets of different publishing departments

As shown in Figure 8 (a), CDC and FDA have the highest level of diffusion and the largest changes in emotion intensity, ranking the top two. HHS and NIH have a small level of diffusion, and their emotion tends to be positive. It can be seen that the tweets of CDC and FDA have a wide range of influence. Among them, CDC is responsible for more specific anti-pandemic management affairs, such as suggestions on community isolation and restrictions on tourism, which are more likely to cause the public to spread negative emotions and finally turn into positive emotions. FDA is responsible for medical support such as diagnosis and treatment technology and drug research and development, and the public once had disputes about its service quality.

As shown in Figure 8 (b), in the process of emotional transmission of tweets published by HHS and NIH, most of the public held positive and neutral emotions. The positive and negative emotions of the public reacting to the tweets issued by CDC and FDA went up and down, and the proportion of negative emotions increased as a whole, while the proportion of positive emotions gradually decreased, and finally the opposite emotions with a very close proportion were formed. In the process of tweets spreading, the FDA timely posted new policies to speed up diagnosis, which promoted the proportion of positive emotions to a certain extent.

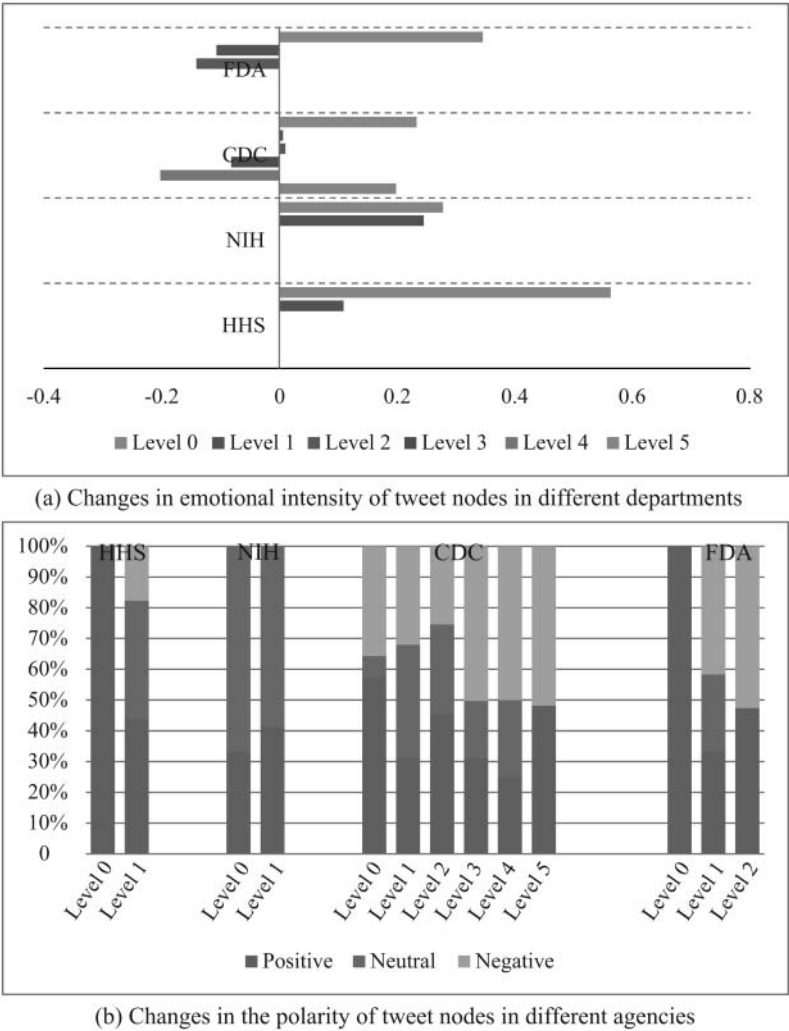


Figure 8. Emotion spread of tweet nodes in different departments.

5. CONCLUSION AND FUTURE WORK

As soon as the tweets related to the pandemic prevention initiative of COVID-19 in the USA were released, the number of tweets directly forwarded accounted for a high proportion in the interaction process at each level. These tweets did not contain the user's comments, so they cannot reflect the real feelings of the users at that time. Therefore, this study ignored the emotion of this part of tweets when analyzing the characteristics of emotional changes. The following four conclusions were drawn in the end. First, the highest level of diffusion in tweets is 6. Second, from the perspective of the time dimension of tweets released, the negative emotions continued to spread among the public in February. In the process of emotional communication of tweets in other months, with the gradual implementation and improvement of the US government's pandemic prevention measures, most of the public emotional diffusion gradually turned from neutral or negative to positive, and the change trend was gradually obvious, especially in May. Thirdly, from the perspective of the topic of tweets, the government's tweets on pandemic related knowledge not only made the public understand the COVID-19 virus scientifically, but also changed their emotions. The public shows more and more negative emotion on the tweets of pandemic prevention management measures, which improved the government's work, and ultimately led to the neutral emotions among the public. The government's alert of pandemic continued to increase public awareness. Fourth, from the perspective of the tweets posters, the tweets issued by CDC and FDA had a wide range of influence, and the public's negative emotions on the specific management affairs and medical support measures of fighting the pandemic in the United States were spread, and finally they tended to account for a very close proportion of the opposite emotions.

In this study, we designed an interaction extraction algorithm of tweets, and proposed a new method to measure the characteristics of the emotional diffusion network in terms of diameter of the network, scope of diffusion, degree of diffusion, and velocity of diffusion, and simultaneously interpreted the characteristics of the emotional diffusion network from two aspects: 1) the intensity of emotional transmission and the change of polarity of emotion, and 2) the influence of key nodes and key users on subsequent emotional intensity. Further research can be done from three aspects. First, the regression analysis of network influencing factors will be more comprehensive, and the trend of emotional diffusion will be predicted. Second, a dynamic analysis system of the emotional diffusion network of tweets will be designed and developed, which can show the process and characteristics of the emotional diffusion network of designated tweets in real time, identify the key nodes and users of emotional diffusion, and predict the trend of emotional diffusion. Finally, the performance of emotion classification will be improved by the supervised learning method and the interaction extraction algorithm of tweets will be optimized by Hadoop cluster to improve the efficiency of the system.

ACKNOWLEDGEMENTS

This work is supported by Humanities and Social Science Research Fund of the Ministry of Education in China (Grant No. 18YJC840045) and Jiangsu Social Science Fund (No. 20TQA001). The authors are grateful to all the anonymous reviewers for their precious comments and suggestions.

AUTHOR CONTRIBUTIONS

This work was a collaboration between all of the authors. H.X. Xi (xihaixu@jsut.edu.cn) conducted the experiment, wrote and revised the paper. C.Z. Zhang (zhangcz@njust.edu.cn) provided the idea of research, designed the experiment and revised the paper. Y. Zhao (zhaoy@njust.edu.cn) visualized the emotional diffusion network. S. He (hs@jsut.edu.cn) preprocessed the data of tweets. All authors have made meaningful and valuable contributions to revising and proofreading the manuscript.

DATA AVAILABILITY STATEMENT

All the data is available in the Science Data Bank repository, <https://doi.org/10.11922/sciencedb.01044>, under an Attribution 4.0 International (CC BY 4.0). To comply with Twitter's Terms of Service, we are only publicly releasing the Tweet IDs of the collected Tweets. The data are released for non-commercial research use.

REFERENCES

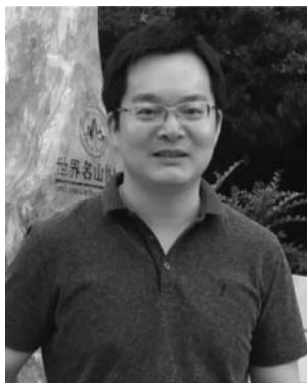
- [1] Raghav Rao, H., et al.: Retweets of officials' alarming vs reassuring messages during the COVID-19 pandemic: Implications for crisis management. *International Journal of Information Management* 55, No. 102187 (2020)
- [2] Prastyo, P., Sumi, A.: Tweets responding to the Indonesian government's handling of COVID-19: Sentiment analysis using SVM with normalized poly Kernel. *Journal of Information Systems Engineering and Business Intelligence* 6, 112–122 (2020)
- [3] Fabiano, N., et al.: An analysis of COVID-19 article dissemination by Twitter compared to citation rates. *medRxiv preprint medRxiv: 2020.06.22.20137505* (2020)
- [4] Ahmed, W., et al.: COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data. *Journal of Medical Internet Research* 22(5), e19458 (2020)
- [5] Park, H.W., Park, S., Chong, M.: Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea. *Journal of Medical Internet Research* 22(5), e18897 (2020)
- [6] Kramer, A.D.I., Guillory J.E., Hancock J.: Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America* 111(24), 8788–90 (2014)
- [7] Li, L., et al.: Effect of anger, anxiety, and sadness on the propagation scale of social media posts after natural disasters. *Information Processing & Management* 57(6), 102313 (2020)
- [8] Zafarani, R., Cole, W.D., Liu, H.: Sentiment propagation in social networks: A case study in LiveJournal. *Advances in Social Computing, Lecture Notes in Computer Science* 6007, 413–420 (2010)
- [9] Peters, K., Kashima, Y.: A multimodal theory of affect diffusion. *Psychological Bulletin* 2015, 141(5), 966–992 (2015)
- [10] Trung, D.N., et al.: Towards modeling fuzzy propagation for sentiment analysis in online social networks: A case study on TweetScope. In: 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), pp. 331–338 (2013)
- [11] Xu, X.: Research on the effect and features of emotional bias in the communication of Sina social news: A Web mining and empirical analysis based on Sina social news (in Chinese). *Chinese Journal of Journalism & Communication* 39(4), 76–94 (2017)

- [12] Miller, M., et al.: Sentiment flow through hyperlink networks. In: International Conference on Weblogs & Social Media, pp. 550–553 (2011)
- [13] Xiong, X., et al.: An emotional contagion model for heterogeneous social media with multiple behaviors. *Physica A Statistical Mechanics & Its Applications* 490, 185–202 (2018)
- [14] Xiong, H., Lv, S.: Factors affecting social media users' emotions regarding food safety issues: Content analysis of a debate among Chinese Weibo users on genetically modified food security. *Healthcare (Basel)* 9(2), No. 113 (2021)
- [15] Institute of Medicine (US) Committee for the Study of the Future of Public Health: The future of public health. National Academies Press (US), Washington (DC) (1988)
- [16] Chen, E., Lerman, K., Ferrara, E.: Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Public Health Surveillance* 6(2), e19273 (2020)
- [17] Hutto, C.J., Gilbert, E.: VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, pp. 216–225 (2014).
- [18] Everett, M., Borgatti, S.: The centrality of groups and classes. *Journal of Mathematical Sociology* 23, 181–201 (1999)

AUTHOR BIOGRAPHY



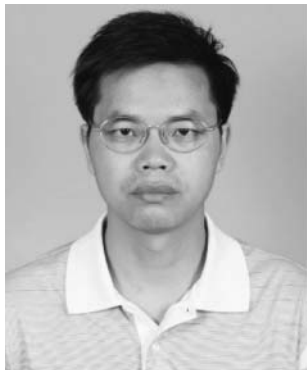
Haixu Xi is currently a doctoral student in the School of Economics and Management, Nanjing University of Science and Technology and an Associate Professor of School of Computer Engineering, Jiangsu University of Technology, China. He received his Master's degree from Nanjing Normal University in 2006. His research interests mainly focus on social media mining, big data application and natural language processing.
ORCID: 0000-0003-4099-9819



Chengzhi Zhang is a Professor of Department of Information Management, Nanjing University of Science and Technology, China. He received his Ph.D. degree of Information Science from Nanjing University, China. He has published more than 100 publications in journals , Including *Journal of the Association for Information Science and Technology* (JASIST), *Aslib Journal of Information Management*, *Journal of Informetrics*, *Online Information Review*, and *Scientometrics*, and conferences, including Annual Meeting of the Association for Computational Linguistics (ACL) and Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). His current research interests include scientific text mining, knowledge entity extraction and evaluation, and social media mining.
ORCID: 0000-0001-9522-2914



Yi Zhao is currently a doctoral student in the School of Economics and Management, Nanjing University of Science and Technology. He received his Master's degree from Hohai University in 2019. His research interests mainly focus on knowledge organization, science of science and natural language processing.
ORCID: 0000-0003-1050-7051



Sheng He is a Professor of School of Computer Engineering, Jiangsu University of Technology, China. He received his B.S. degree in Physics from Fuyang Normal University in 1993 and M.E. degree in Software Engineering from University of Science and Technology of China in 2005. His research interests include data mining and big data application.

ORCID: 0000-0001-6762-8271